Barcelona School of Economics

RESEARCH NOTE

An exploration of psychometric networks

Niamh Sheehan and Benjamin Seimon

Introduction

[1] Modelling psychological phenomena as causal systems or networks is increasing in popularity. The nascent field of network psychometrics has the potential to make important contributions to psychopathology (Borsboom, Cramer, et al. 2013). Unlike typical networks studied in fields such as economics, psychological networks are networks between variables, not entities. Nodes do not represent individuals, banks, or proteins but rather observed psychological variables. Often the variables are responses to questions and can cover mood states, symptoms, or attitudes. Edges in this context indicate the presence of a statistical relationship between nodes.

This conceptualisation of psychological experiences challenges the prevailing common cause hypothesis or reflective model in psychopathology (Borsboom, Cramer, et al. 2013). Rather than being caused by an unobserved latent factor such as a disorder, a network-based approach assumes variables interact directly with each other (Marian, Sava, and Dindelegan 2022). In other words, symptoms are not indicators of a common cause (e.g. PTSD) but rather co-occur in syndromes due to causal interactions (modelled by the network).

The latest techniques for network estimation are particularly pertinent in this field since knowing the 'true structure' of a psychological model is challenging without sufficiently formalised theory (Van Borkulo et al. 2014). Network models have been proven to highlight dynamics that might lead to an individual developing or continuing to suffer from a mental illness (Epskamp et al. 2016). The ability to visualise the multivariate dependencies may reveal dynamics that would remain unknown otherwise. The use of network measures can support the identification of central symptoms to the causal system. This could support decision makers in a clinical setting to prioritise treatments (Marian, Sava, and Dindelegan 2022). Applications of network analysis for psychometrics are briefly summarised in the table below:

The following paper begins by replicating figures from Armour et al. (2017) to demonstrate a cross-sectional data application. We then explore the use of vector autoregressive models (VAR) to estimate and analyse temporal and contemporaneous networks for time-series data. We focus on investigating

1. We acknowledge having heavily relied on (Epskamp, 2016) for the foundations of this paper and proving a wonderful introduction into the topic of network psychometrics.

Table 1. Model Details Question 2

Data Type	Method	Output
Cross Sectional	Gaussian graphical model	Network of partial correlation coefficients
(Multi) single-subject time series	*(Multi-level)* vector autoregression	Temporal Network Contemporaneous Network *Between-subjects network*

the relationships between two nodes via edge weight, the importance of individual nodes through different measures of node centrality (namely betweenness and closeness), and understanding the relationships between variables via community detection.

Application 1: Cross-sectional data

Cross-sectional analysis involves applying a model to a dataset which measures variables related to multiple individuals at a single time stamp. To estimate network models pairwise Markov random fields (PMRF) are most commonly used. Other popular methods are Gaussian graphical models (GGM). A GGM network is based on partial correlation coefficients, the correlation between two variables after conditioning on all other variables present in the dataset. When estimating partial correlation networks, each link represents a partial correlation coefficient (ranging from –1 to 1) between two nodes. (Epskamp et al. 2016). One such example is demonstrated in Figures 1 and 2.

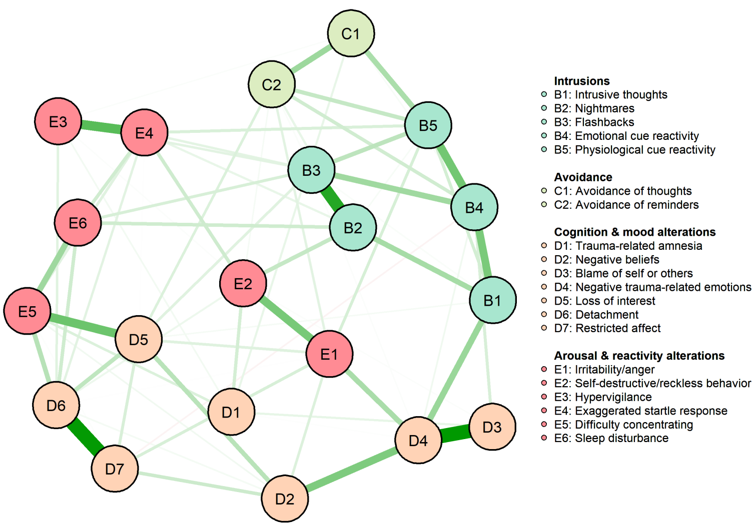

Figure 1. Estimated GGM network for 20 PTSD symptoms

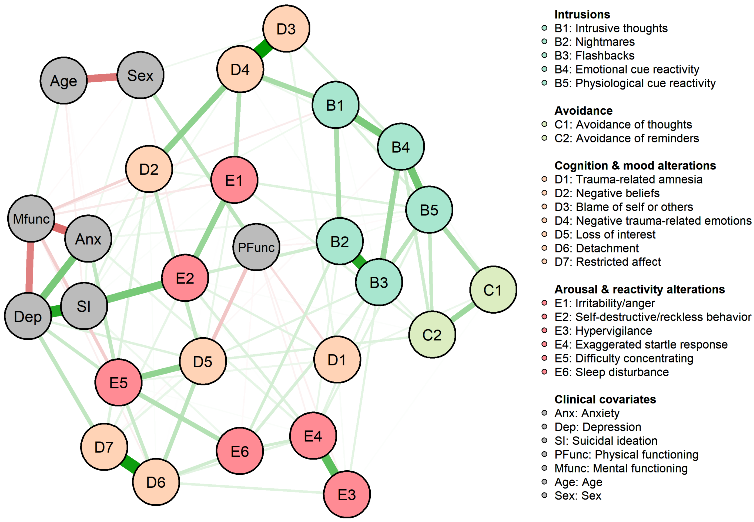

Figure 2. Estimated GGM network for 20 PTSD symptoms and 7 clinical covariates

Figure 1 shows a network containing the twenty symptoms of PTSD according to the DSM-5 framework [2]. Positive (negative) associations are in green (red). The thickness and brightness of an edge indicate strength of association. The nodes within the network are mostly positively correlated with the strongest connection observed between the two avoidance variables (PCC of .59). By contrast the strongest negative correlation is between D7 and D4 (PCC of -0.054). Figure 2 is a network which contains the twenty symptoms along with seven clinical covariates. Although many of the connections remain similar to the network in Figure 1, some relationships in this network are not explained by covariation among the twenty PTSD symptoms alone. For example, self-destructive behaviour and suicidal ideation (E2 and SI) are positively correlated while negative beliefs and mental quality of life (D2 and Mfunc) are negatively correlated.

Of high interest are the instances where the partial correlation is exactly zero. That is, when the two variables are conditionally independent. Due to sampling variation none of the correlation coefficients are exactly zero. This is problematic because as a result even conditionally independent nodes have a weak edge between them. These edges are considered to be spurious. Regularisation within LASSO estimation is the most popular method to address this issue. When the hyperparameter (lambda) is sufficiently high, all connections are removed. Similarly, when lambda equals zero or is low enough, all connections remain. As a result, the careful selection of lambda is crucial. However, after experimentation with lambda we opted for the EBIC penalty parameter which performs better through lower false positive rates (Van Borkulo et al. 2014; Chen and Chen 2008; Barber and Drton 2015). Here, the gamma parameter decides the strength of the penalty on including additional neighbours. When gamma is zero no penalty is given for the number of neighbours which naturally results in a higher number of estimated connections. Equally, with gamma set to one, a high penalty is applied to greater neighbours. Tuning of this parameter in the R *qgraph* package is unnecessary [3].

2. DSM-5-TR is the standard classification of mental disorders used by mental health professionals in the United States.
3. The glasso is run for 100 values of the parameter logarithmically spaced between maximal value of the tuning parameter at which all edges are zero, the maximum lambda and the maximum lambda/100. For each graph the EBIC is computed and

Finally, the most common approach for considering the importance of relationships within the network is to report node centrality measures and edge weights. Again, we show this for the same dataset in Figure 3 and 4. As is evident from the plots, the estimates are strikingly linked. Negative trauma-related emotions (D4) is the most central variable across all measures in the network which is intuitive given that the syndrome in question is PTSD. The centrality of other variables is less consistent across the measure of centrality.

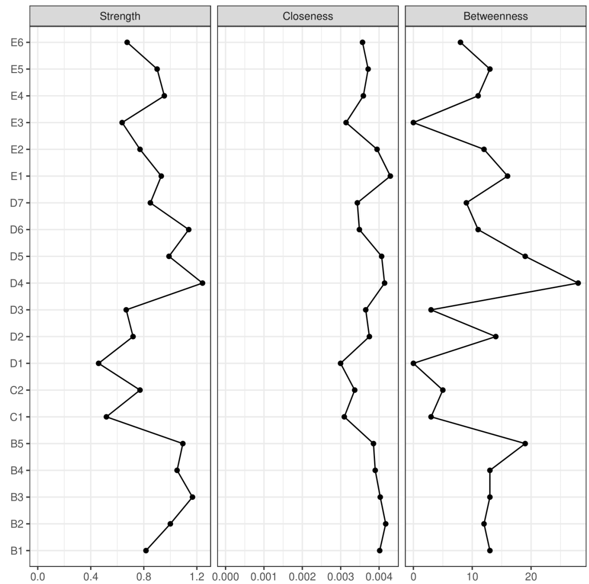

Figure 3. Centrality measures: betweenness, closeness, and node strength centrality estimates for the 20 DSM-5 PTSD criterion symptoms.

the best is returned

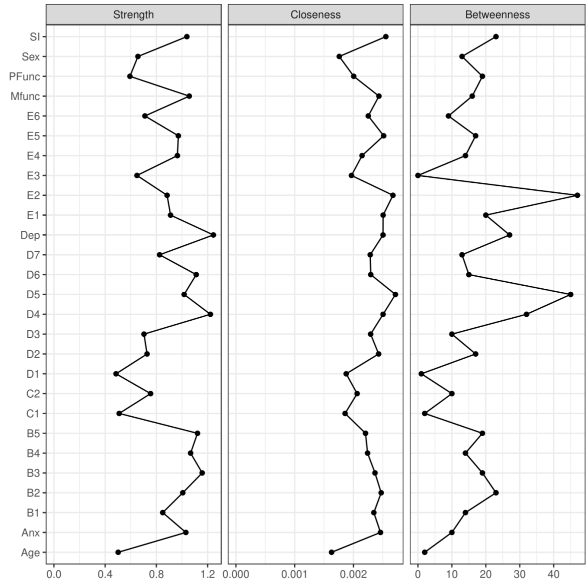

Figure 4. Centrality measures: betweenness, closeness, and node strength centrality estimates for the 20 DSM-5 PTSD criterion symptoms and 7 clinical covariates

Application 2: Single Subject Time-Series

1. Motivation

Recent technological advances have now enabled researchers within psychopathology to collect time-intensive, repeated measurements. This time-series data is most often sourced from experience sampling method (ESM) studies (Epskamp et al. 2016). Typically, individuals will respond to a set of questions relating to their symptoms at several regular intervals during the day (termed the 'window of measurement') over a period of weeks or months. Considering the relationship between nodes over time is a valuable exercise, since it allows for the investigation of the dynamic relationship between symptoms and other variables for a single or group of individual/s.

This section of the paper details the application of statistical techniques from network psychometrics to a single-subject time-series dataset. First, a brief overview of the theoretical underpinnings of VARs is presented. Then the *graphicalVAR* package from R is implemented to estimate both a temporal and contemporaneous network before a high-level analysis of the results.

2. Theory

It is critical to note that from a statistical perspective the assumption of independent observations is clearly violated when working with ESM studies (e.g. someone who is tired between 0900 and 1100 is still likely to be tired between 1100 and 1300). Hence, network analysis predominantly involves the use of VAR models to deal with autocorrelation. This requires making some strict assumptions in order for the model to be fully specified (Epskamp et al. 2016), which we detail below for the n=1 case:

1. The joint probability distribution can be factorised according to a graph.
2. The conditional probability distributions are stable and independent of t (stationarity).
3. The first measurements are treated as exogenous and not modelled.
4. The conditional distributions are multivariate normal.

Intuitively, these models suggest that a given variable at t can be predicted by the same variable at *t-1* (autoregressive effects) and all other variables at *t-1* (cross-lagged effects). The model is flexible and can be specified to include additional lags, but the focus of this paper is on a lag 1 model. As portrayed in Table 1, statistical tools can be used to estimate three different types of networks from time series data. The between-subjects network, which is an undirected partial correlation network between the means of the different individuals' scores, is beyond the scope of this paper [4]. Instead this paper focuses on:

1. **The temporal network**: a directed network detailing relationships between variables across time.
2. **The contemporaneous network**: an undirected partial correlation network detailing relationships between variables within a single window of measurement.

The temporal network can cautiously be interpreted as being a causal representation of the dynamics of the system since the cause, by definition, must precede the effect. As such, this network is directed. However, residuals of the temporal VAR model are also correlated. The residuals enable relationships *between* windows of measurement and *within* windows of measurement to be separated. In other words there are correlations at *t* that cannot be explained by the temporal effects, which can be used to compute a network of partial correlations - the contemporaneous network.

Before detailing the application to a 'real-world' dataset, it is important to first discuss whether these networks can be used to infer causality. Given the nature of psychological disorders, there is some debate amongst scholars as to the usefulness of temporal networks (Epskamp et al. 2016). Although measurements are taken regularly, it is still likely insufficient to fully characterise the dynamic interplay between symptoms. For example, the time between experiencing the effects of precursors to a panic attack (e.g. heart palpitations) and the anticipation of a panic attack is likely to be far shorter (minutes) than the window of measurement (typically hours).

Hence, the contemporaneous network perhaps provides greater scope for inferring causal relationships since it captures correlations *within* a unit of measurement. Recalling that causality is reflected within the structure of conditional independence, the value of a partial correlation network is clear. However, since the graph is undirected, generating 'causal hypotheses' is required. This requires both domain knowledge, and an understanding of the individual context in order to glean insight that may be helpful within clinical practice. Furthermore, descriptive measures of centrality and edge weight are useful for interpretation.

3. Data

The data for the single-subject time series application is obtained from (Bringmann et al. 2021). A customised questionnaire was designed for a 31 year old woman suffering from OCD with comorbid depressive symptoms. Both positive and negative variables are captured by the questionnaire, which was completed as long as the participant deemed it useful. In total, the dataset contains responses

4. Only applicable for multi-subject data.

from 15/03/2017 to 30/04/2018 across 27 variables. In theory, a complete dataset would include three measurements (9am, 12pm and 9pm) for every day. This would enable the use of two lags – a 'day' lag as well as a 'measurement time' lag. However, as expected with a self-driven survey, the dataset contains a large number of missing values:

1. **Incomplete observations**: for a survey completed on a given day and time, a response across all 27 variables is frequently not present. We choose to only include observations that have a complete response by dropping variables with an excess of 1000 missing values (Qualitysleep, Pleasant, Unpleasant and Remarks). Binary variables are dropped since their inclusion led to spurious correlations.
2. **Gaps between and within days**: After taking the above steps, there are significant gaps between recordings – the maximum gap is 6 days. Furthermore, on any given day, there is extensive missingness with respect to a morning, afternoon, and evening response. This presents a clear challenge given that we propose a lag 1 model. Including all the data would undermine the robustness of the analysis since we would consider the codependence between two measurements on day t as equivalent to a measurement at 9:00 on day t and 21:00 on day t+1. Therefore, we choose to subset the data to the first observation recorded on a given day. This allows us to account for the between-day dependencies of observations.

The final dataset contains 329 observations of 21 variables. The scoring is on a scale from zero to one hundred.

4. Methodology and Results

Given the structure of our revised dataset, we choose a one lag model as follows:

$$y_t = \beta y_{t-1} + e_t$$

Errors are assumed to be distributed normally with mean vector 0, an inverse covariance matrix Kappa and are independent *between* time points but not *within* time points. The matrix of Beta and Kappa coefficients encodes *between* and *within* time point interactions respectively. Two separate tuning parameters are therefore required in the implementation of the LASSO in order to control the level of sparsity in Beta and Kappa. The model is estimated under a 50x50 grid of tuning parameters, and the optimal choice is decided according to the EBIC criterion [5]. The Beta and Kappa matrices are then standardised to give:

- Partial directed correlations: $PCC(y_{i,t}, y_{j,t}) = \frac{\kappa_{i,j}}{\sqrt{\kappa_{i,i}\kappa_{j,j}}}$
- Partial contemporaneous correlations: $PDC(y_{i,t}, y_{j,t}) = \frac{\beta_{i,j}}{\sqrt{\sigma_{j,j}\kappa_{i,i}+\beta^2_{j,i}}}$

Figure 5 and 6 present the estimated partial contemporaneous correlation network and partial directed correlation network.

5. The minimum tuning parameters for Beta and Kappa are set to 0.05 as per Epskamp et al. (2016)

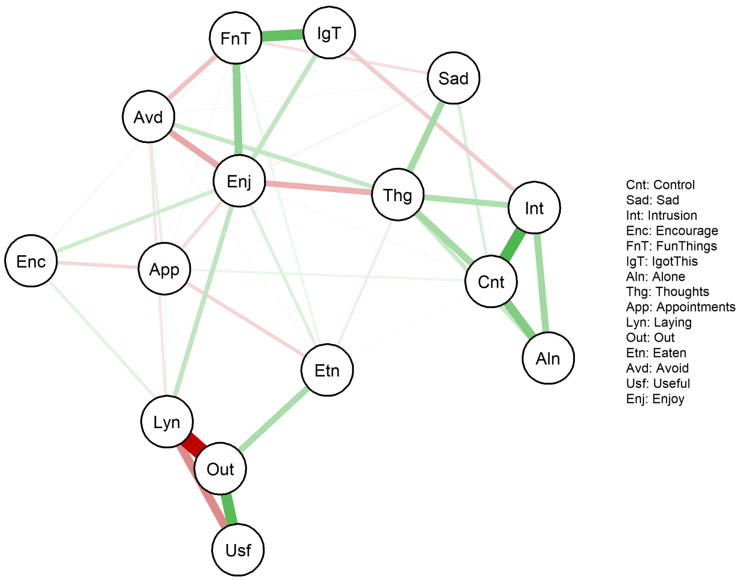

Figure 5. Estimated PCC network from ESM data

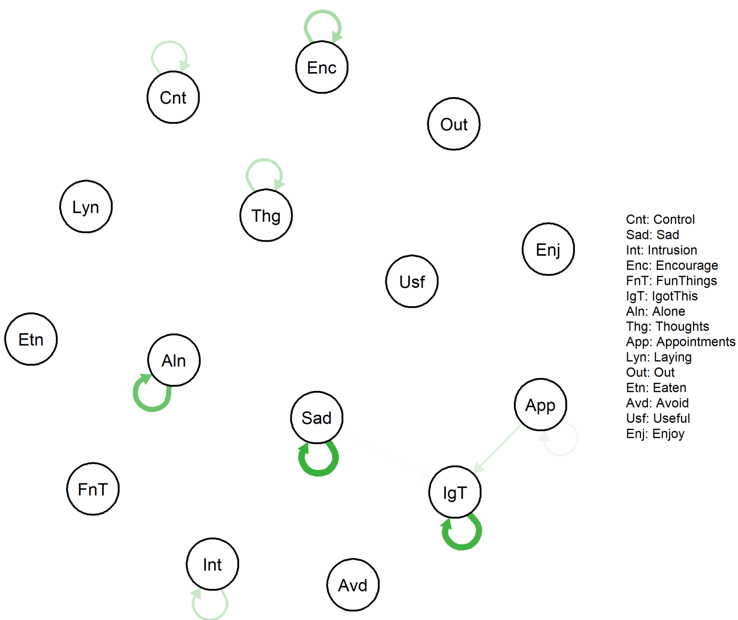

Figure 6. Estimated PDC network from ESM data

The partial directed correlations show strong self-loops between 'Igotthis' (0.47), 'Sad' (0.46) and 'Alone' (0.39). This is in line with logical reasoning, where we would not expect self loops between variables such as "Eaten". This also highlights the usefulness of this analysis for exploring thought patterns, since for example we may be more likely to observe a self-loop for "Eaten" for an individual suffering from an eating disorder.

The contemporaneous network shows a range of positive and negative partial correlations. The strongest positive correlations are observed beteween "Intrusion" and "Control", "IgotThis and "FunThings" as well as "Out" and "Useful". In contrast, the strongest negative correlations are observed between "Outside" and "Laying" as well as "Laying" and "Useful". This could be useful from a treatment perspective, where avoiding periods of 'laying' could support positive mental wellbeing.

We are encouraged by our results given that less than 10 per cent of the computed PCC's are above 0 and below 0.1. This shows that the sparse estimation has succesfully removed spurious edges, whilst retaining edges that are likely to be 'true' representations of the network.

Next, we consider the centrality measures (see Figure 7). Across all measures Enjoy is the found to be the most central. Encourage is the least central. Unlike in the cross-sectional case the values are consistent regardless of the measure chosen, but the betweenness displays higher variation that strength and closeness.

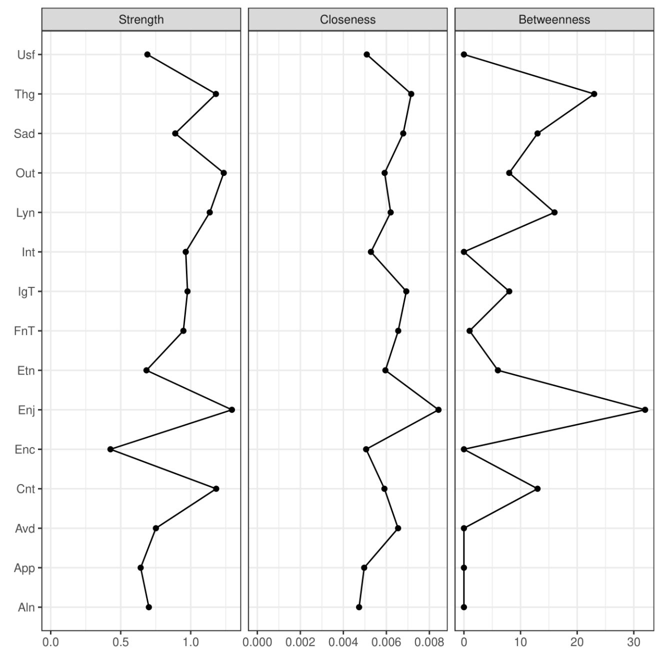

Figure 7. Centrality Measures: betweenness, closeness, and node strength estimates for the ESM Data

Given the strength of partial correlations between subsets of nodes, we apply the Louvain community detection algorithm to the contemporaneous network as a further extension. In short, the algorithm:

1. Assigns all nodes to their own community
2. Selects the node i to move into community j which maximises the modularity [6] score.
3. Stops when there are no moves which could increase modularity.
4. Nodes are then grouped into their respective community and the procedure is repeated with the new network structure

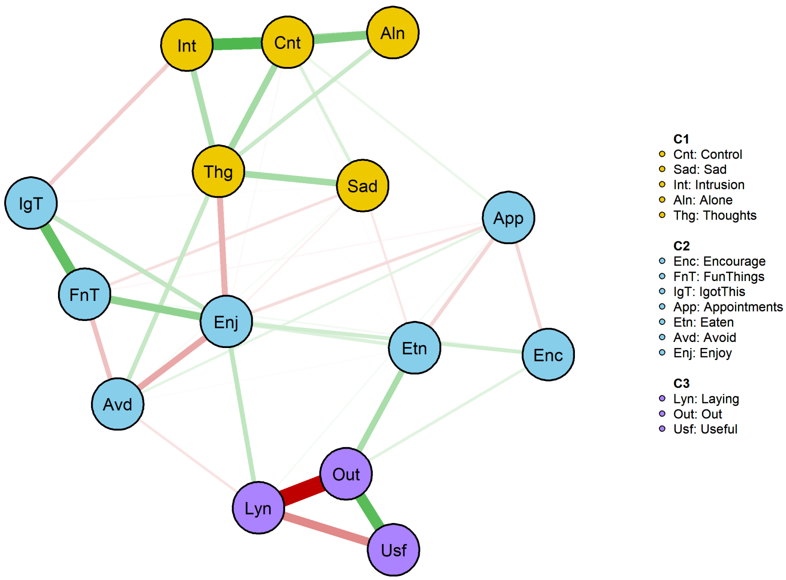

Figure 8. Figure 9: Estimated PCC network from ESM data with Louvain community detection

The algorithm results in a modularity score of 0.39 and a visual interpretation of the communities is shown in Figure 8. We see that the community detection algorithm performs well, recognising the strength of the relationships observed in the PCC in Figure 5.

Conclusion

There are two primary shortcomings of the analysis presented above. The first is the difficulty in assessing the networks' accuracy. When the true network structure is known, specificity and sensitivity [7] can be used to evaluate the performance of the estimation. However in psychological networks this is not possible. Researchers most often use non-parametric bootstrapping to assess network accuracy. Bootstrapped difference tests determine whether edge-weights and centrality significantly differ across bootstrapped samples. This is often presented with a 95% confidence interval. Similarly case-dropping subset bootstrapping for stability appraisal are prevalent in the

6. Modularity measures systematic deviations from a random configuration for each possible partition. This is the prevailing metric for community detection, whereby a 'good' partition separates groups into those with higher internal connections than would be expected at random.

7. Sensitivity expresses the proportion of true connections which are correctly estimated as present, and is also known as the true positive rate. Specificity corresponds to the proportion of absent connections which are correctly estimated as zero, and is also known as the true negative rate. (nature)

literature (Epskamp et al. 2016). Given the exploratory nature of this paper we did not present a bootstrapped version of the measures but bootstrapping could be used to test the robustness of our results.

As outlined in the data description for the second application, the ESM dataset suffers from severe missingness. Even when subsetting to the first survey response of each day, the distribution of time measurements is 141 in the morning, 120 in the afternoon and 68 in the evening. However, 133 of the 329 are recorded at the same time as the previous day, 154 are recorded at plus or minus one time interval, and only 41 are plus or minus 2. We accept that this is a limitation of the study, but note that there are limited alternatives given the nature of the dataset. For example, one could use forward or backward fills, but we judged this to be equally spurious.

Despite these limitations, this paper clearly establishes the potential role of network analysis in the field of psychology. Whether the data is cross sectional or time series, suitable methods exist with which we can further understand the relationships between various symptoms and clinical variables. Through two separate applications we demonstrate the ease and viability of constructing networks using partial correlation coefficients and offer various means of analysing the corresponding networks.

References

Armour, Cherie, Eiko I Fried, Marie K Deserno, Jack Tsai, and Robert H Pietrzak. 2017. A network analysis of dsm-5 posttraumatic stress disorder symptoms and correlates in us military veterans. *Journal of anxiety disorders* 45:49–59.

Barber, Rina Foygel, and Mathias Drton. 2015. High-dimensional ising model selection with bayesian information criteria. *Electronic Journal of Statistics* 9 (1): 567–607.

Borsboom, Denny, Angélique OJ Cramer, et al. 2013. Network analysis: an integrative approach to the structure of psychopathology. *Annual review of clinical psychology* 9 (1): 91–121.

Bringmann, Laura F, Date C van der Veen, Marieke Wichers, Harriëtte Riese, and Gert Stulp. 2021. Esmvis: a tool for visualizing individual experience sampling method (esm) data. *Quality of Life Research* 30 (11): 3179–3188.

Chen, Jiahua, and Zehua Chen. 2008. Extended bayesian information criteria for model selection with large model spaces. *Biometrika* 95 (3): 759–771.

Epskamp, Sacha, Gunter KJ Maris, Lourens J Waldorp, and Denny Borsboom. 2016. Network psychometrics. *arXiv preprint arXiv:1609.02818.*

Marian, Ştefan, Florin A Sava, and Camelia Dindelegan. 2022. A network analysis of dsm-5 avoidant personality disorder diagnostic criteria. *Personality and Individual Differences* 188:111454.

Van Borkulo, Claudia D, Denny Borsboom, Sacha Epskamp, Tessa F Blanken, Lynn Boschloo, Robert A Schoevers, and Lourens J Waldorp. 2014. A new method for constructing networks from binary data. *Scientific reports* 4 (1): 1–10.

5. Appendix

Link to Github

A variety of specialist R packages were used to conduct the analysis presented:

1. qgraph
2. graphicalVAR
3. bootnet
4. IsingFit
5. mgm

Scoring is done on a scale with a slider from 0 to 100 (unless indicated otherwise)

The first set of questions concerns momentary assessments

1. I am afraid to lose control [Scale = (not at all – very much)]
2. I feel sad / useless / meaningless [Scale = (not at all – very much)]
3. How convincing are the intrusions [Scale = (not at all – very much)]
4. I can encourage myself [Scale = (not at all – very much)]
5. I feel like doing something fun [Scale = (not at all – very much)]
6. I have the feeling that "I can do this" [Scale = (not at all – very much)]
7. Are you in company? [Yes / No]
8. How afraid are you of being alone? [Scale = (not at all – very much)]

The next set of questions concerns occurrences since the last measurement point

9. How often have you been in contact with someone you feel safe with? [Score: 0–1–2–3–4]
10. How often have you thought of contacting someone you feel safe with? [Scale = (not at all – very often)]
11. How strong was your inclination to cancel appointments? [Scale = (not at all – very much)]
12. Have you actually cancelled appointments? [Yes / No]
13. Have you laid on the couch or in bed since the last measurement point? [Scale = (not at all – very often)]
14. Have you slept since the last measurement point? [Yes / No]
15. If yes: How did you sleep? [Scale = (badly – very well)]
16. Did you leave the house? [Scale = (not at all – very often)]
17. How did you eat? [Scale = (not at all – very well)]
18. Have you avoided everyday things? [Scale = (not at all – very much)]
19. Have you done useful (important) things? [Scale = (not at all – very much)]
20. Have you enjoyed your activities? [Scale = (not at all – very much)]
21. Since the last measurement point, have you experienced any (un)pleasant everyday occurrences?
22. Yes, something pleasant: How pleasant was this experience? [Scale = (not at all – very much)]
23. Yes, something unpleasant: How unpleasant was this experience? [Scale = (not at all – very much)]

If you want, you can add comments here:

[space for comments]

Concluding sentence:

'Thank you very much for filling out the questionnaire, and do not forget to charge the battery of your smartphone.'

Figure 9. Questionnaire from (Bringmann et al. 2021)